

FPTAS for Hardcore and Ising Models on Hypergraphs

Pinyan Lu*

Kuan Yang[†]Chihao Zhang[‡]

Abstract

Hardcore and Ising models are two most important families of two state spin systems in statistic physics. Partition function of spin systems is the center concept in statistic physics which connects microscopic particles and their interactions with their macroscopic and statistical properties of materials such as energy, entropy, ferromagnetism, etc. If each local interaction of the system involves only two particles, the system can be described by a graph. In this case, fully polynomial-time approximation scheme (FPTAS) for computing the partition function of both hardcore and anti-ferromagnetic Ising model was designed up to the uniqueness condition of the system. These result are the best possible since approximately computing the partition function beyond this threshold is NP-hard. In this paper, we generalize these results to general physics systems, where each local interaction may involves multiple particles. Such systems are described by hypergraphs. For hardcore model, we also provide FPTAS up to the uniqueness condition, and for anti-ferromagnetic Ising model, we obtain FPTAS where a slightly stronger condition holds.

*Microsoft Research. pinyanl@microsoft.com

[†]Shanghai Jiao Tong University. peanutyk@gmail.com

[‡]Shanghai Jiao Tong University. chihao.zhang@gmail.com

1 Introduction

In recent couple of years, there are remarkable progress on designing approximate counting algorithms based on correlation decay approach [Wei06, BG08, GK12, RST⁺11, LLY12, SST12, LLY13, LY13, LLL14, LWZ14, LLZ14, LL15b, LL15a]. Unlike the previous major approximate counting approach that based on random sampling such as Markov Chain Monte Carlo (MCMC) (see for examples [JS97, JS93, JSV04, GJ11, DJV02, Jer95, Vig99, DFJ02, DG00, LV97]), correlation decay based approach provides deterministic fully polynomial-time approximation scheme (FPTAS). New FPTASes were designed for a number of interesting combinatorial counting problems and computing partition functions for statistic physics systems, where partition function is a weighted counting function from the computational point of view. One most successful example is the algorithm for anti-ferromagnetic two-spin systems [LLY12, SST12, LLY13], including counting independent sets [Wei06]. The correlation decay based FPTAS is beyond the best known MCMC based FPRAS and achieves the boundary of approximability [SS12, GSV12].

In this paper, we generalize these results of anti-ferromagnetic two-spin systems to hypergraphs. For physics point of view, this corresponds to spin systems with higher order interactions, where each local interaction involves more than two particles. There are two main ingredients for the original algorithms and analysis on normal graphs (we will use the term normal graph for a graph to emphasize that it is not hypergraphs): (1) the construction of the self-avoiding walk tree by Weitz [Wei06], which transform a general graph to a tree; (2) correlation decay proof for the tree, which enables one to truncate the tree to get a good approximation in polynomial time. However, the construction of the self-avoiding walk tree cannot be extended to hypergraphs, which is the main obstacle for the generalization.

The most related previous work is counting independent sets for hypergraphs by Liu and Lu [LL15b]. They established a computation tree with a two-layers recursive function instead of the self-avoiding walk tree and provided a FPTAS to count the number of independent sets for hypergraphs with maximum degree of 5, extending the algorithm for normal graph with the same degree bound. Their proof was significantly more complicated than the previous one due to the complication of the two-layers recursive function. In particular, the “right” degree bound for the problem is a real number between 5 and 6 if one allow fraction degree in some sense. This integer gap provides some room of flexibility and enables them to do some case-by-case numerical argument to complete the proof. However, the parameters for the anti-ferromagnetic two-spin systems on hypergraphs are real numbers. To get a sharp threshold, we do not have any room for numerical approximation.

1.1 Our results

We study two most important anti-ferromagnetic two-spin systems on hypergraphs: the hardcore model and the anti-ferromagnetic Ising model. The formal definitions of these two models can be found in Section 2.

Our first result is an FPTAS to compute the partition function of hypergraph hardcore model.

Theorem 1. *For hardcore model with a constant activity parameter of λ , there is an FPTAS to compute the partition function for hypergraphs with maximum degree $\Delta \geq 2$ if $\lambda < \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$.*

This bound is exactly the uniqueness threshold for the hardcore on normal graphs. Thus, it is tight since normal graphs are special cases of hypergraphs. To approximately compute the partition function beyond this threshold is NP-hard. In particular, The FPTAS in [LL15b] for counting the number of independent sets for hypergraphs with maximum degree of 5 can be viewed

as a special case of our result with parameters $\Delta = 5, \lambda = 1$, which satisfies the above uniqueness condition. Another interesting special case is when $\Delta = 2$. This is not an interesting case for normal graphs since a normal graph with maximum degree of 2 is simply a disjoint union of paths and cycles, whose partition function can be computed exactly. However, the problem becomes more complicated on hypergraphs: it can be interpreted as counting weighted edge covers on normal graphs by viewing vertices of degree two as edges and hyperedges as vertices. The exact counting of this problem is known to be $\#P$ -complete and an FPTAS was found recently [LLZ14]. In our model, the uniqueness bound $\frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ is infinite for $\Delta = 2$ and as a result we give an FPTAS for counting weighted edge covers for any constant edge weight λ . This gives an alternative proof for the main result in [LLZ14].

Our second result is on computing the partition function of anti-ferromagnetic Ising model.

Theorem 2. *For Ising model with interaction parameter $0 < \beta < 1$ and external field λ , there is an FPTAS to compute the partition function for hypergraphs with maximum degree Δ if $\beta \geq 1 - \frac{2}{2e^{-1/2}\Delta+3}$.*

The tight uniqueness bound for anti-ferromagnetic Ising model on normal graphs is $\beta \geq 1 - \frac{2}{\Delta}$. So, our bound is in the same asymptotic order but a bit worse in the constant coefficient as $2e^{-1/2} \approx 1.213 > 1$. Moreover, our result can apply beyond Ising model to a larger family of anti-ferromagnetic two-spin systems on hypergraphs.

1.2 Our techniques

We also use the correlation decay approach. Although the framework of this method is standard, in many work along this line of research, new tools and techniques are developed to make this relatively new approach more powerful and widely applicable. This is indeed the case for the current paper as well. We summarize the new techniques we introduced here.

For hardcore model, we replace the numerical case-by-case analysis by a monotone argument with respect to the edge size of the hypergraph which shows that the normal graphs with edge size of 2 is indeed the worst cases. This gives a tight bound for hardcore model.

To handle hypergraph with unbounded edge sizes, we need to prove that the decay rate is much smaller for edges of larger size. Such effect is called computationally efficient correlation decay, which has been used in many previous works to obtain FPTASes for systems with unbounded degrees or edge sizes. In all those works, one sets a threshold for the parameter and proves different types of bounds for large and small ones separately. Such artificial separation gets a discontinuous bound which adds some complications in the proof and usually ends with a case-by-case discussion. In particular, this separation is not compatible with the above monotone argument. To overcome this, we propose a new uniform and smooth treatment for this by modifying the decay rate by a polynomial function of the edge size. After this modification, we only need to prove one single bound which automatically provides computationally efficient correlation. We believe that this idea is important and may find applications in other related problems.

For the Ising model, the main difficulty is to get a computation tree as a replacement of the self-avoiding walk tree. We proposed one, which also works for general anti-ferromagnetic two-spin systems on hypergraphs. However, unlike the case of the hardcore model, the computation tree is not of perfect efficiency and this is the main reason that the bound we achieve in Theorem 2 is not tight. To get the computationally efficient correlation decay, we also use the above mentioned uniform and smooth treatment. We also extend our result beyond Ising to a family of anti-ferromagnetic two-spin systems on hypergraphs.

1.3 Discussion and open problems

One obvious open question is to close the gap for Ising model, or more generally extend our work to anti-ferromagnetic two-spin systems on hypergraphs with better parameters. However, it seems that it is impossible to obtain a tight result in these models using the computation tree proposed in this paper, due to its imperfectness. How to overcome this is an important open question.

Even for the hardcore model, our result is tight only for the family of all hypergraphs since the normal graphs are special cases. From both physics and combinatorics point of view, it would be very interesting to study the family of w -uniform hypergraphs where each hyperedge is of the same size w . By our monotone argument, it is plausible to conjecture that one can get better bound for larger w . In particular, MCMC based approach does show that larger edge size helps: for hypergraph independent set with maximum degree of Δ and minimum edge size w , an FPRAS for $w \geq 2\Delta + 1$ was shown in [BDK08]. However, their result is not tight. Can we get a tight bounds in terms of Δ and w by correlation decay approach? The high level idea sounds promising, but there is an obstacle to prove such result by our computation tree. To construct the computation tree, we need to construct modified instances. In these modified instances, the size of a hyperedge may decrease to as low as 2. Therefore, even if we start with w -uniform hypergraphs or hypergraphs with minimum edge size of w , we may need to handle the worst case of normal graphs during the analysis. How to avoid this effect is a major open question whose solution may have applications in many other problems.

The fact that larger hyperedge size only makes the problem easier is not universally true for approximation counting. One interesting example is counting hypergraph matchings. FPTAS for counting 3D matchings of hypergraphs with maximum degree 4 is given in [LL15b], and extension to weighted setting are studied in [YZ15]. In particular, a uniqueness condition in this setting is defined in [YZ15], and it is a very interesting open question whether this uniqueness condition is also the transition boundary for approximability.

2 Preliminaries

A hypergraph $G(V, \mathcal{E})$ consists of a vertex set V and a set of hyperedges $\mathcal{E} \subseteq 2^V$. For every hyperedge $e \in \mathcal{E}$ and vertex $v \in V$, we use $e - v$ to denote $e \setminus \{v\}$ and use $e + v$ to denote $e \cup \{v\}$.

2.1 Hypergraph hardcore model

The hardcore model is parameterized by the activity parameter $\lambda > 0$. Let $G(V, \mathcal{E})$ be a hypergraph. An independent set of G is a vertex set $I \subseteq V$ such that $e \not\subseteq I$ for every hyperedge $e \in \mathcal{E}$. We use $\mathcal{I}(G)$ to denote the set of independent sets of G . The weight of an independent set I is defined as $w(I) \triangleq \lambda^{|I|}$. We let $Z(G)$ denote the partition function of $G(V, \mathcal{E})$ in the hardcore model, which is defined as

$$Z(G) \triangleq \sum_{I \in \mathcal{I}(G)} w(I).$$

The weight of independent sets induces a Gibbs measure on G . For every $I \in \mathcal{I}$, we use

$$\Pr_G[I] \triangleq \frac{w(I)}{Z(G)}$$

to denote the probability of obtaining I if we sample according to the Gibbs measure. For every $v \in V$, we use

$$\Pr_G[v \in I] \triangleq \sum_{\substack{I \in \mathcal{I}(G) \\ v \in I}} \Pr_G[I]$$

to denote the *marginal probability* of v .

2.2 Hypergraph two state spin model

Now we give a formal definition to hypergraph two state spin systems. This model is parameterized by the external field $\lambda > 0$. An instance of the model is a labeled hypergraph $G(V, \mathcal{E}, (\beta, \gamma))$ where $\beta, \gamma : \mathcal{E} \rightarrow \mathbb{R}$ are two labeling functions that assign each edge $e \in \mathcal{E}$ two reals $\beta(e), \gamma(e)$. A configuration on G is an assignment $\sigma : V \rightarrow \{0, 1\}$ whose weight $w(\sigma)$ is defined as

$$w(\sigma) \triangleq \prod_{e \in \mathcal{E}} w(e, \sigma) \prod_{v \in V} w(v, \sigma)$$

where for a hyperedge $e = \{v_1, \dots, v_w\}$

$$w(e, \sigma) \triangleq \begin{cases} \beta(e) & \text{if } \sigma(v_1) = \sigma(v_2) = \dots = \sigma(v_w) = 0 \\ \gamma(e) & \text{if } \sigma(v_1) = \sigma(v_2) = \dots = \sigma(v_w) = 1 \\ 1 & \text{otherwise} \end{cases}$$

and for a vertex v ,

$$w(v, \sigma) \triangleq \begin{cases} \lambda & \text{if } \sigma(v) = 1 \\ 1 & \text{otherwise.} \end{cases}$$

The partition function of the instance is given by

$$Z(G) = \sum_{\sigma \in \{0,1\}^V} w(\sigma).$$

Similarly, the weight of configurations induces a Gibbs measure on G . For every $\sigma \in \{0, 1\}^V$, we use

$$\Pr_G[\sigma] \triangleq \frac{w(\sigma)}{Z(G)}$$

to denote the probability of σ in the measure. For every $v \in V$, we use

$$\Pr_G[\sigma(v) = 1] \triangleq \sum_{\substack{\sigma \in \{0,1\}^V \\ \sigma(v)=1}} \Pr_G[\sigma]$$

to denote the *marginal probability* of v .

The *anti-ferromagnetic* Ising model is the special case that $\beta \triangleq \beta(e) = \gamma(e) \leq 1$ for all $e \in \mathcal{E}$. In this model, we call β the *interaction parameter* of the model. The hardcore model introduced in previous section is the special case that $\beta(e) = 1$ and $\gamma(e) = 0$ for all $e \in \mathcal{E}$.

We give the whole proof to Theorem 2 in appendix. More precisely, we design an FPTAS for the more general two state spin system and establish the following theorem:

Theorem 3. Consider a class of two state spin system with external field λ such that each instance $G(V, \mathcal{E}, (\beta, \gamma))$ in the class satisfies $1 - \frac{2}{2e^{-1/2\Delta} + 3} \leq \beta(e), \gamma(e) \leq 1$ where Δ is the maximum degree of G . There exists an FPTAS to compute the partition function for every instance in the class.

Theorem 2 then follows since it is a special case of Theorem 3.

Actually, the main idea of FPTAS design and proof for this model is similar to the idea we use to solve hypergraph hardcore model. However, the details of recursion function design and techniques for proof of correlation decay property are pretty different from that in hypergraph hardcore model, so we put the whole section in appendix.

3 Hypergraph Hardcore Model

3.1 Recursion for computing marginal probability

We first fix some notations on graph modification specific to hypergraph independent set. Let $G(V, \mathcal{E})$ be a hypergraph.

- For every $v \in V$, we denote $G - v \triangleq (V \setminus \{v\}, \mathcal{E}')$ where $\mathcal{E}' \triangleq \{e \setminus \{v\} \mid e \in \mathcal{E}\}$.
- For every $e \in \mathcal{E}$, we denote $G - e \triangleq (V, \mathcal{E} \setminus \{e\})$.
- Let x be a vertex or an edge and y be a vertex or an edge, we denote $G - x - y \triangleq (G - x) - y$.
- Let $S = \{v_1, \dots, v_k\} \subseteq V$, we denote $G - S \triangleq G - v_1 - v_2 \dots - v_k$.
- Let $\mathcal{F} = \{e_1, \dots, e_k\} \subseteq \mathcal{E}$, we denote $G - \mathcal{F} \triangleq G - e_1 - e_2 \dots - e_k$.

Let $G(V, \mathcal{E})$ be a hypergraph and $v \in V$ be an arbitrary vertex with degree d . Let $\{e_1, \dots, e_d\}$ be the set of hyperedges incident to v and for every $i \in [d]$, $e_i = \{v\} \cup \{v_{ij} \mid j \in [w_i]\}$ consists of $w_i + 1$ vertices.

We first define a graph $G'(V', \mathcal{E}')$, which is the graph obtained from G by replacing v by d copies of itself and each e_i contains a distinct copy. Formally, $V' \triangleq (V \setminus \{v\}) \cup \{v_1, \dots, v_d\}$, $\mathcal{E}' \triangleq \{e \in \mathcal{E} \mid v \notin e\} \cup \{e_i - v + v_i \mid i \in [d]\}$.

For every $i \in [d]$ and $j \in [w_i]$, we define a hypergraph $G_{ij}(V_{ij}, \mathcal{E}_{ij})$:

$$G_{ij} \triangleq G' - \{v_k \mid i \leq k \leq d\} - \{e_k \mid 1 \leq k \leq i\} - \{v_{ik} \mid 1 \leq k < j\}.$$

Let $R_v = \frac{\Pr_G[v \in I]}{\Pr_G[v \notin I]}$ and $R_{ij} = \frac{\Pr_{G_{ij}}[v_{ij} \in I]}{\Pr_{G_{ij}}[v_{ij} \notin I]}$. We can compute R_v by following recursion:

Lemma 4.

$$R_v = \lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} \frac{R_{ij}}{1 + R_{ij}} \right). \quad (1)$$

The proof of this lemma is postponed to appendix.

The Uniqueness Condition Let the underlying graph be an infinite d -ary tree, then the recursion (1) becomes

$$f_{\lambda, d}(x) = \lambda \left(\frac{1}{1 + x} \right)^d.$$

Let \hat{x} be the positive fixed-point of $f_{\lambda, d}(x)$, i.e., $\hat{x} > 0$ and $f_{\lambda, d}(\hat{x}) = \hat{x}$. The condition on λ for the uniqueness of the Gibbs measure is that $|f'_{\lambda, d}(\hat{x})| < 1$. The following proposition is well-known.

Proposition 5. Let $\lambda_c = \frac{d^d}{(d-1)^{d+1}}$, then $\left| f'_{\lambda_c, d}(\hat{x}) \right| = 1$ and for every $0 < \lambda < \lambda_c$, it holds that $\left| f'_{\lambda, d}(\hat{x}) \right| < 1$.

3.2 The algorithm to compute marginal probability

Let $G(V, \mathcal{E})$ be a hypergraph with maximum degree Δ and $v \in V$ be an arbitrary vertex with degree d . Define G_{ij} , R_v , R_{ij} as in Section 3.1. Then the recursion (1) gives a way to compute the marginal probability $\Pr_G[v \in I]$ *exactly*. However, an exact evaluation of the recursion requires a computation tree with exponential size. Thus we introduce the following truncated version of the recursion, with respect to constants $c > 0$ and $0 < \alpha < 1$.

$$R(G, v, L) = \begin{cases} \lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} \frac{R(G_{ij}, v_{ij}, L)}{1 + R(G_{ij}, v_{ij}, L)} \right) & \text{if } d = \Delta \\ \lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} \frac{R(G_{ij}, v_{ij}, L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor)}{1 + R(G_{ij}, v_{ij}, L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor)} \right) & \text{if } d < \Delta \text{ and } L > 0 \\ \lambda & \text{otherwise.} \end{cases}$$

The recursion can be directly used to compute $R(G, v, L)$ for any given L and it induces a truncated computation tree (with height L in some special metric). It is worth noting that, the case that $d = \Delta$ can only happen at the root of the computation tree, since in each smaller instance, the degree of v_{ij} is decreased by at least one.

We claim that $R(G, v, L)$ is a good estimate of R_v with a suitable choice of c and α , for those (λ, Δ) in the uniqueness region.

Lemma 6. Let $G(V, \mathcal{E})$ be a hypergraph with maximum degree $\Delta \geq 2$. Let $v \in V$ be a vertex with degree d and let $\lambda < \lambda_c = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ be the activity parameter. There exist constants $C > 0$ (more precisely, $C = 6\lambda\sqrt{1+\lambda}$) and $\alpha \in (0, 1)$ such that

$$|R(G, v, L) - R_v| \leq C \cdot \alpha^{\max\{0, L\}}$$

for every L .

The whole proof of this lemma is postponed to the next section.

Proof of Theorem 1. Assuming Lemma 6, the proof of Theorem 1 is routine and we put the proof into appendix. \square

3.3 Correlation decay

In this section, we establish Lemma 6. We first prove some technical lemmas.

Suppose $f : D^d \rightarrow \mathbb{R}$ is a d -ary function where $D \subseteq \mathbb{R}$ is a convex set, let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be an increasing differentiable function and $\Phi(x) \triangleq \phi'(x)$. The following proposition is a consequence of the mean value theorem:

Proposition 7. For every $\mathbf{x} = (x_1, \dots, x_d)$, $\hat{\mathbf{x}} = (\hat{x}_1, \dots, \hat{x}_d) \in D^d$, it holds that

1. $|f(\mathbf{x}) - f(\hat{\mathbf{x}})| = \frac{1}{\Phi(\tilde{x})} |\phi(f(\mathbf{x})) - \phi(f(\hat{\mathbf{x}}))|$ for some $\tilde{x} \in D$;
2. $|\phi(f(\mathbf{x})) - \phi(f(\hat{\mathbf{x}}))| \leq \sum_{i=1}^d \frac{\Phi(f)}{\Phi(\tilde{x}_i)} \left| \frac{\partial f(\tilde{\mathbf{x}})}{\partial x_i} \right| \cdot |\phi(x_i) - \phi(\hat{x}_i)|$ for some $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_d) \in D^d$.

Lemma 8. Let $\Delta \geq 2$ be a constant integer and $\lambda < \lambda_c = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ be a constant real. Let $d < \Delta$ and $w_1, \dots, w_d > 0$ be integers and $f = \lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} \frac{x_{ij}}{1+x_{ij}}\right)$ be a $\left(\sum_{i=1}^d w_i\right)$ -ary function. Let $\Phi(x) = \frac{1}{\sqrt{x(1+x)}}$. Let $c < \min \left\{ \frac{\log(1+\lambda) - \log \lambda}{2+4\lambda}, \frac{2\lambda+1}{2} \log \left(\frac{1+\lambda}{\lambda} \right) - 1 \right\}$ be a positive number. There exists a constant $\alpha < 1$ depending on λ and d (but not depending on w_i for all $i \in [d]$) such that

$$\sum_{a=1}^d w_a^c \sum_{b=1}^{w_a} \frac{\Phi(f)}{\Phi(x_{ab})} \left| \frac{\partial f(\mathbf{x})}{\partial x_{ab}} \right| \leq \alpha < 1$$

for every $\mathbf{x} = (x_{ij})_{i \in [d], j \in [w_i]}$ where each $x_{ij} \in [0, \lambda]$

The lemma bounds the amortized decay rate, which is the key to the proof of correlation decay. In previous works, the amortized decay rate is defined as

$$\sum_{a=1}^d \sum_{b=1}^{w_a} \frac{\Phi(f)}{\Phi(x_{ab})} \left| \frac{\partial f(\mathbf{x})}{\partial x_{ab}} \right|,$$

without the w_a^c factor. Then one need to give a constant $\alpha < 1$ bound for small w_a and a sub constant bound for large w_a . With this modification, we only need to prove a single bound as above.

Notice that we require c to be a positive constant, so it is necessary to verify that $\frac{2\lambda+1}{2} \log \left(\frac{1+\lambda}{\lambda} \right) - 1 > 0$ for every $\lambda > 0$. To see this, let $h(\lambda) \triangleq \frac{2\lambda+1}{2} \log \left(\frac{1+\lambda}{\lambda} \right) - 1$, then we can compute that

$$\begin{aligned} h'(\lambda) &= \log \left(\frac{1+\lambda}{\lambda} \right) - \frac{1+2\lambda}{2\lambda+2\lambda^2}, \\ h''(\lambda) &= \frac{1}{2\lambda^2(1+\lambda)^2}. \end{aligned}$$

Since $h''(\lambda) > 0$ for every λ , $h'(\lambda)$ is increasing. Along with the fact that $\lim_{\lambda \rightarrow \infty} h'(\lambda) = 0$, we have $h'(\lambda) < 0$ for every $\lambda > 0$. This implies that $h(\lambda)$ is decreasing. Also note that

$$\lim_{\lambda \rightarrow \infty} h(\lambda) = \lim_{\lambda \rightarrow \infty} \log \left(\left(1 + \frac{1}{\lambda}\right)^\lambda \left(1 + \frac{1}{\lambda}\right)^{1/2} \right) - 1 = 0.$$

It holds that $h(\lambda) > 0$ for every $\lambda > 0$. Thus a positive c satisfying $c < h(\lambda)$ exists for every $\lambda > 0$.

Proof of Lemma 8. To simplify the notation, we first let $t_{ij} = \frac{x_{ij}}{1+x_{ij}}$, then for every $i \in [d]$ and $j \in [w_i]$, it holds that $t_{ij} \in \left[0, \frac{\lambda}{1+\lambda}\right]$ and

$$f = \lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} t_{ij}\right).$$

For every $a \in [d]$ and $b \in [w_a]$, we have

$$\left| \frac{\partial f}{\partial x_{ab}} \right| = \lambda (1 - t_{ab})^2 \prod_{\substack{j \in [w_a] \\ j \neq b}} t_{aj} \cdot \prod_{\substack{i \in [d] \\ i \neq a}} \left(1 - \prod_{j=1}^{w_i} t_{ij}\right) = f \cdot \frac{(1 - t_{ab})^2}{t_{ab}} \cdot \frac{\prod_{j=1}^{w_a} t_{aj}}{1 - \prod_{j=1}^{w_a} t_{aj}}.$$

Thus

$$\sum_{a=1}^d w_a^c \sum_{b=1}^{w_a} \frac{\Phi(f)}{\Phi(x_{ab})} \left| \frac{\partial f}{\partial x_{ab}} \right| = \sqrt{\frac{f}{1+f}} \sum_{a=1}^d \frac{w_a^c \prod_{j=1}^{w_a} t_{aj}}{1 - \prod_{j=1}^{w_a} t_{aj}} \sum_{b=1}^{w_a} \frac{1 - t_{ab}}{\sqrt{t_{ab}}}.$$

Let $\mathbf{t} = (t_{ij})_{i \in [d], j \in [w_i]}$, define

$$\begin{aligned} h(\mathbf{t}) &\triangleq \sqrt{\frac{f}{1+f}} \sum_{a=1}^d \frac{w_a^c \prod_{j=1}^{w_a} t_{aj}}{1 - \prod_{j=1}^{w_a} t_{aj}} \sum_{b=1}^{w_a} \frac{1 - t_{ab}}{\sqrt{t_{ab}}} \\ &= \sqrt{\frac{\lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} t_{ij}\right)}{1 + \lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} t_{ij}\right)}} \sum_{a=1}^d \frac{w_a^c \prod_{j=1}^{w_a} t_{aj}}{1 - \prod_{j=1}^{w_a} t_{aj}} \sum_{b=1}^{w_a} \frac{1 - t_{ab}}{\sqrt{t_{ab}}}. \end{aligned}$$

For every $\mathbf{t} = (t_{ij})_{i \in [d], j \in [w_i]}$ where each $t_{ij} \in [0, \frac{\lambda}{1+\lambda}]$, define a tuple $\hat{\mathbf{t}} = (\hat{t}_{ij})_{i \in [d], j \in [w_i]}$ such that for every $i \in [d]$,

$$\hat{t}_{ij} = \begin{cases} \left(\frac{1+\lambda}{\lambda}\right)^{w_i-1} \prod_{k=1}^{w_i} t_{ik} & \text{if } j = 1 \\ \frac{\lambda}{1+\lambda} & \text{otherwise.} \end{cases}$$

We claim that $h(\mathbf{t}) \leq h(\hat{\mathbf{t}})$. To see this, first note that for every $i \in [d]$, $\prod_{j=1}^{w_i} t_{ij} = \prod_{j=1}^{w_i} \hat{t}_{ij}$, it is sufficient to prove that for every $i \in [d]$

$$\sum_{j=1}^{w_i} \frac{1 - t_{ij}}{\sqrt{t_{ij}}} \leq \sum_{j=1}^{w_i} \frac{1 - \hat{t}_{ij}}{\sqrt{\hat{t}_{ij}}}.$$

This is a consequence of the Karamata's inequality by noticing that the function $\frac{1-e^x}{\sqrt{e^x}}$ is convex.

We rename \hat{t}_{i1} to t_i and it is sufficient to upper bound

$$g(\mathbf{t}, \mathbf{w}) \triangleq \sqrt{\frac{\lambda \prod_{i=1}^d \left(1 - \left(\frac{\lambda}{1+\lambda}\right)^{w_i-1} t_i\right)}{1 + \lambda \prod_{i=1}^d \left(1 - \left(\frac{\lambda}{1+\lambda}\right)^{w_i-1} t_i\right)}} \cdot \sum_{i=1}^d \frac{w_i^c \left(\frac{\lambda}{1+\lambda}\right)^{w_i-1} t_i}{1 - \left(\frac{\lambda}{1+\lambda}\right)^{w_i-1} t_i} \cdot \left(\frac{1 - t_i}{\sqrt{t_i}} + \frac{(w_i - 1)}{\sqrt{\lambda + \lambda^2}}\right) \quad (2)$$

where $t_i \in [0, \frac{\lambda}{1+\lambda}]$ and $w_i \in \mathbb{Z}^+$ for every $i \in [d]$.

The argument so far is similar to the proof in [LL15b]. In the following, we prove a monotonicity property of each w_i and thus avoid the heavy numerical analysis in [LL15b] and allow us to obtain a tight result.

For every $i \in [d]$, we let $z_i \triangleq 1 - \left(\frac{\lambda}{1+\lambda}\right)^{w_i-1} t_i$ and thus equivalently $t_i = (1 - z_i) \left(\frac{1+\lambda}{\lambda}\right)^{w_i-1}$. For every fixed $\mathbf{z} = (z_1, \dots, z_d)$, we can write (2) as

$$g_{\mathbf{z}}(\mathbf{w}) = \sqrt{\frac{\lambda \prod_{i=1}^d z_i}{1 + \lambda \prod_{i=1}^d z_i}} \sum_{i=1}^d \frac{1 - z_i}{z_i} \left(\frac{1 - t_i}{\sqrt{t_i}} + \frac{w_i - 1}{\sqrt{\lambda + \lambda^2}}\right) w_i^c. \quad (3)$$

We show that $g_{\mathbf{z}}(\mathbf{w})$ is monotonically decreasing with w_i for every $i \in [d]$.

Denote $T_i \triangleq \frac{1-t_i}{\sqrt{t_i}} + \frac{(w_i-1)}{\sqrt{\lambda+\lambda^2}}$, then

$$\frac{\partial g_{\mathbf{z}}(\mathbf{w})}{\partial w_i} = \sqrt{\frac{\lambda \prod_{i=1}^d z_i}{1 + \lambda \prod_{i=1}^d z_i}} \cdot \frac{1 - z_i}{z_i} \left(\frac{\partial T_i}{\partial w_i} w_i^c + c w_i^{c-1} T_i\right). \quad (4)$$

The partial derivative (4) is negative for a suitable choice of c :

$$\begin{aligned}
& \frac{1-z_i}{z_i} \cdot \left(\frac{\partial T_i}{\partial z_i} w_i^c + c w_i^{c-1} T_i \right) \\
&= \frac{1-z_i}{z_i} \cdot \left(\left(-\frac{1}{2} t_i' (t_i^{-1/2} + t_i^{-3/2}) + \frac{1}{\sqrt{\lambda + \lambda^2}} \right) w_i^c + c w_i^{c-1} \left(\frac{1-t_i}{\sqrt{t_i}} + \frac{(w_i-1)}{\sqrt{\lambda + \lambda^2}} \right) \right) \\
&= \frac{1-z_i}{z_i} \cdot w_i^{c-1} \left(\left(-\frac{1}{2} \log \left(\frac{1+\lambda}{\lambda} \right) (t_i^{1/2} + t_i^{-1/2}) + \frac{1}{\sqrt{\lambda + \lambda^2}} \right) w_i + c \left(t_i^{-1/2} - t_i^{1/2} + \frac{(w_i-1)}{\sqrt{\lambda + \lambda^2}} \right) \right) \\
&= \frac{1-z_i}{z_i} \cdot w_i^{c-1} \left(\frac{(c+1)w_i - c}{\sqrt{\lambda + \lambda^2}} - \left(t_i^{1/2} \left(\frac{1}{2} w_i \log \left(\frac{1+\lambda}{\lambda} \right) + c \right) + t_i^{-1/2} \left(\frac{1}{2} w_i \log \left(\frac{1+\lambda}{\lambda} \right) - c \right) \right) \right)
\end{aligned}$$

Denote

$$p(t, w) \triangleq \frac{(c+1)w - c}{\sqrt{\lambda + \lambda^2}} - \left(t^{1/2} \left(\frac{1}{2} w \log \left(\frac{1+\lambda}{\lambda} \right) + c \right) + t^{-1/2} \left(\frac{1}{2} w \log \left(\frac{1+\lambda}{\lambda} \right) - c \right) \right)$$

Since $c \leq \frac{\log(1+\lambda) - \log \lambda}{2+4\lambda}$, the term

$$t^{1/2} \left(\frac{1}{2} w \log \left(\frac{1+\lambda}{\lambda} \right) + c \right) + t^{-1/2} \left(\frac{1}{2} w \log \left(\frac{1+\lambda}{\lambda} \right) - c \right)$$

achieves its minimum at $t = \frac{\lambda}{1+\lambda}$. Thus

$$p(t, w) \leq p \left(\frac{\lambda}{1+\lambda}, w \right) = \left(\frac{\lambda}{1+\lambda} \right)^{1/2} \left(\frac{c+1}{\lambda} - \frac{2\lambda+1}{2\lambda} \log \left(\frac{1+\lambda}{\lambda} \right) \right) w.$$

Moreover, $c < \frac{2\lambda+1}{2} \log \left(\frac{1+\lambda}{\lambda} \right) - 1$ implies that $\frac{c+1}{\lambda} < \frac{2\lambda+1}{2\lambda} \log \left(\frac{1+\lambda}{\lambda} \right)$ holds, which consequently leads to $p \left(\frac{\lambda}{1+\lambda}, 1 \right) < 0$.

In all, we choose a positive constant $c < \min \left\{ \frac{\log(1+\lambda) - \log \lambda}{2+4\lambda}, \frac{2\lambda+1}{2} \log \left(\frac{1+\lambda}{\lambda} \right) - 1 \right\}$, and this results in $p \left(\frac{\lambda}{1+\lambda}, w \right) \leq p \left(\frac{\lambda}{1+\lambda}, 1 \right) < 0$.

In light of the monotonicity of w_i 's, for every fixed \mathbf{z} , $g_{\mathbf{z}}(\mathbf{w})$ achieves its maximum when $\mathbf{w} = \mathbf{1}$. Thus

$$\begin{aligned}
\max_{\mathbf{t} \in [0, \frac{\lambda}{1+\lambda}]^d} g(\mathbf{t}, \mathbf{w}) &= \max_{\substack{\mathbf{z}=(z_1, \dots, z_d) \\ \forall i \in [d], z_i \in [1 - (\frac{\lambda}{1+\lambda})^{w_i-1}, 1]}} g_{\mathbf{z}}(\mathbf{w}) \leq \max_{\substack{\mathbf{z}=(z_1, \dots, z_d) \\ \forall i \in [d], z_i \in [1 - (\frac{\lambda}{1+\lambda})^{w_i-1}, 1]}} g_{\mathbf{z}}(\mathbf{1}) \leq \max_{\mathbf{z} \in [0, 1]^d} g_{\mathbf{z}}(\mathbf{1}).
\end{aligned}$$

Actually, the case that all w_i 's are 1 corresponds to counting weighted independent sets on normal graphs and arguments to bound $g_{\mathbf{z}}(\mathbf{1})$ can be found in [LLY13]. For the sake of completeness, we give a proof of $g_{\mathbf{z}}(\mathbf{1}) \leq \alpha < 1$ (Lemma 9) in appendix. \square

Lemma 9. *Let $\Delta > 1$, be a constant. Assume $\lambda < \lambda_c = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ be a constant and $d < \Delta$. Then for some constant $\alpha < 1$, $g_{\mathbf{z}}(\mathbf{1}) \leq \alpha < 1$ where $\mathbf{z} = (z_1, \dots, z_d) \in [0, 1]^d$.*

We are now ready to prove the main lemma.

Proof of Lemma 6. Let $\Phi(x) = \frac{1}{\sqrt{x(1+x)}}$ and $\phi(x) = \int \Phi(x) dx = 2 \sinh^{-1}(\sqrt{x})$. We first apply induction on $\ell \triangleq \max \{0, L\}$ to show that, if $d < \Delta$, then $|\phi(R_v) - \phi(R(G, v, L))| \leq 2\sqrt{\lambda}\alpha^L$ for some constant $\alpha < 1$.

If $\ell = 0$, note that $R_v \in [0, \lambda]$, thus $|\phi(R(G, v, L)) - \phi(R_v)| \leq 2\sqrt{\lambda}$. We now assume $L = \ell > 0$ and the lemma holds for smaller ℓ . For every $i \in [d]$ and $j \in [w_i]$, we denote $x_{ij} = R_{ij}$ and $\hat{x}_{ij} = R(G_{ij}, v_{ij}, L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor)$. Let $\mathbf{x} = (x_{ij})_{i \in [d], j \in [w_i]}$, $\tilde{\mathbf{x}} = (\hat{x}_{ij})_{i \in [d], j \in [w_i]}$. Let $f = \lambda \prod_{i=1}^d \left(1 - \prod_{j=1}^{w_i} \frac{x_{ij}}{1-x_{ij}}\right)$, then it follows from Proposition 7 that for some $\tilde{\mathbf{x}} = (\tilde{x}_{ij})_{i \in [d], j \in [w_i]}$ with each $\tilde{x}_{ij} \in [0, \lambda]$

$$\begin{aligned} |\phi(R_v) - \phi(R(G, v, L))| &\leq \sum_{i=1}^d \sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(\tilde{x}_{ij})} \left| \frac{\partial f(\tilde{\mathbf{x}})}{\partial x_{ij}} \right| \cdot |\phi(x_{ij}) - \phi(\hat{x}_{ij})| \\ &\stackrel{(\spadesuit)}{\leq} 2\sqrt{\lambda} \sum_{i=1}^d \sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(\tilde{x}_{ij})} \left| \frac{\partial f(\tilde{\mathbf{x}})}{\partial x_{ij}} \right| \alpha^{L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor} \\ &\stackrel{(\heartsuit)}{\leq} 2\sqrt{\lambda} \alpha^L. \end{aligned}$$

(\spadesuit) follows from the induction hypothesis and (\heartsuit) is due to Lemma 8.

The case that $d = \Delta$ can only happen at the root of our computational tree. Following the arguments in the proofs of 8, 9 and the bound in (5), it is easy to see that a universal constant upper bound for the error contraction exists, i.e.,

$$\sum_{i=1}^{\Delta} \sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(\tilde{x}_{ij})} \left| \frac{\partial f(\tilde{\mathbf{x}})}{\partial x_{ij}} \right| < \max_{z \in [0,1]} \sqrt{\frac{\Delta^2(\Delta-1)^{\Delta-1} z^{\Delta}(1-z)}{(\Delta-2)^{\Delta} + (\Delta-1)^{\Delta-1} z^{\Delta}}} < 3.$$

Thus $|\phi(R_v) - \phi(R(G, v, L))| \leq 6\sqrt{\lambda} \alpha^L$ for every v .

Then the lemma follows from Proposition 7, since

$$\begin{aligned} |R_v - R(G, v, L)| &= \frac{1}{\Phi(\tilde{x})} \cdot |\phi(R_v) - \phi(R(G, v, L))| \quad \text{for some } \tilde{x} \in [0, \lambda] \\ &\leq 6\lambda\sqrt{1+\lambda} \cdot \alpha^L \end{aligned}$$

□

References

- [BDK08] Magnus Bordewich, Martin Dyer, and Marek Karpinski. Path coupling using stopping times and counting independent sets and colorings in hypergraphs. *Random Structures & Algorithms*, 32(3):375–399, 2008.
- [BG08] Antar Bandyopadhyay and David Gamarnik. Counting without sampling: Asymptotics of the log-partition function for certain statistical physics models. *Random Structures & Algorithms*, 33(4):452–479, 2008.
- [DFJ02] Martin E. Dyer, Alan M. Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. *SIAM Journal on Computing*, 31(5):1527–1541, 2002.
- [DG00] Martin E. Dyer and Catherine S. Greenhill. On Markov Chains for independent sets. *Journal of Algorithms*, 35(1):17–49, 2000.
- [DJV02] Martin Dyer, Mark Jerrum, and Eric Vigoda. Rapidly mixing Markov chains for dismantlable constraint graphs. In *Randomization and Approximation Techniques in Computer Science*, pages 68–77. Springer, 2002.

- [GJ11] Leslie Ann Goldberg and Mark Jerrum. A polynomial-time algorithm for estimating the partition function of the ferromagnetic ising model on a regular matroid. In *Proceedings of ICALP*, pages 521–532, 2011.
- [GK12] David Gamarnik and Dmitriy Katz. Correlation decay and deterministic FPTAS for counting colorings of a graph. *Journal of Discrete Algorithms*, 12:29–47, 2012.
- [GSV12] A. Galanis, D. Stefankovic, and E. Vigoda. Inapproximability of the partition function for the antiferromagnetic ising and hard-core models. *Arxiv preprint arXiv:1203.2226*, 2012.
- [Jer95] Mark Jerrum. A very simple algorithm for estimating the number of k -colorings of a low-degree graph. *Random Structures & Algorithms*, 7(2):157–166, 1995.
- [JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the ising model. *SIAM Journal on Computing*, 22(5):1087–1116, 1993.
- [JS97] Mark Jerrum and Alistair Sinclair. *The Markov chain Monte Carlo method: an approach to approximate counting and integration*, pages 482–520. PWS Publishing Co., Boston, MA, USA, 1997.
- [JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *Journal of the ACM*, 51:671–697, July 2004.
- [LL15a] Jingcheng Liu and Pinyan Lu. FPTAS for $\#BIS$ with degree bounds on one side. In *Proceedings of STOC (to appear)*, 2015.
- [LL15b] Jingcheng Liu and Pinyan Lu. FPTAS for counting monotone CNF. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1531–1548, 2015.
- [LLL14] Chengyu Lin, Jingcheng Liu, and Pinyan Lu. A simple FPTAS for counting edge covers. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 341–348, 2014.
- [LLY12] Liang Li, Pinyan Lu, and Yitong Yin. Approximate counting via correlation decay in spin systems. In *Proceedings of SODA*, pages 922–940, 2012.
- [LLY13] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *Proceedings of SODA*, pages 67–84, 2013.
- [LLZ14] Jingcheng Liu, Pinyan Lu, and Chihao Zhang. FPTAS for counting weighted edge covers. In *Algorithms - ESA 2014 - 22th Annual European Symposium, Wroclaw, Poland, September 8-10, 2014. Proceedings*, pages 654–665, 2014.
- [LV97] Michael Luby and Eric Vigoda. Approximately counting up to four (extended abstract). In *Proceedings of STOC*, pages 682–687, 1997.
- [LWZ14] Pinyan Lu, Menghui Wang, and Chihao Zhang. FPTAS for weighted fibonacci gates and its applications. In *ICALP (1)*, pages 787–799, 2014.

- [LY13] Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 16th International Workshop, APPROX 2013, and 17th International Workshop, RANDOM 2013, Berkeley, CA, USA, August 21-23, 2013. Proceedings*, pages 639–654, 2013.
- [RST⁺11] Ricardo Restrepo, Jinwoo Shin, Prasad Tetali, Eric Vigoda, and Linji Yang. Improved mixing condition on the grid for counting and sampling independent sets. In *Proceedings of FOCS*, pages 140–149, 2011.
- [SS12] Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d -regular graphs. In *Proceedings of FOCS*, pages 361–369, 2012.
- [SST12] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. In *Proceedings of SODA*, pages 941–953, 2012.
- [Vig99] Eric Vigoda. Improved bounds for sampling coloring. In *Proceedings of FOCS*, pages 51–59, 1999.
- [Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of STOC*, pages 140–149, 2006.
- [YZ15] Yitong Yin and Jinman Zhao. Counting hypergraph matchings up to uniqueness threshold. *arXiv preprint arXiv:1503.05812*, 2015.

A Proof of Theorem 1

Proof. The input of the FPTAS is an instance $G(V, \mathcal{E})$ and an accuracy parameter $0 < \varepsilon < 1/2$. Assume $V = \{v_1, \dots, v_n\}$. Note that $I = \emptyset$ is an independent set of G with $w(I) = 1$. Therefore

$$Z(G) = 1/\Pr_G[I] = \left(\Pr_G \left[\bigwedge_{i=1}^n v_i \notin I \right] \right)^{-1} = \left(\prod_{i=1}^n \Pr_G \left[v_i \notin I \mid \bigwedge_{j=1}^{i-1} v_j \notin I \right] \right)^{-1}.$$

For every $1 \leq i \leq n$, we define a graph $G_i(V_i, E_i)$:

- $G_1 \triangleq G$;
- For every $i \geq 2$, $G_i \triangleq G_{i-1} - v_{i-1} - \mathcal{E}'$ where $\mathcal{E}' \triangleq \{e \in \mathcal{E}_{i-1} \mid v_{i-1} \in e\}$ consists of edges in G_{i-1} incident to v_i .

It is straightforward to verify that $\Pr_G[v_i \notin I \mid \bigwedge_{j=1}^{i-1} v_j \notin I] = \Pr_{G_i}[v_i \notin I]$ for every $1 \leq i \leq n$. Thus,

$$Z(G) = \prod_{i=1}^n (\Pr_{G_i}[v_i \notin I])^{-1} = \prod_{i=1}^n (1 + R_i),$$

where $R_i \triangleq \frac{\Pr_{G_i}[v_i \in I]}{\Pr_{G_i}[v_i \notin I]}$. Let C and α be constants in Lemma 6. We compute $R(G_i, v_i, L)$ with $L = \frac{\log(2Cn\varepsilon^{-1})}{\log \alpha^{-1}}$ for every $1 \leq i \leq n$, then

$$|R_i - R(G_i, v_i, L)| \leq \frac{\varepsilon}{2n}.$$

This implies

$$1 - \frac{\varepsilon}{2n} \leq \frac{1 + R_i}{1 + R(G, v_i, L)} \leq 1 + \frac{\varepsilon}{2n}.$$

Let $\hat{Z} = \prod_{i=1}^n (1 + R(G_i, v_i, L))^{-1}$ be our estimate of the partition function, then it holds that

$$e^{-\varepsilon} \leq \frac{Z(G)}{\hat{Z}} \leq e^{\varepsilon}.$$

It remains to bound the running time of our algorithm. Let $T(L)$ denote the maximum running time of computing $R(G, v, L)$ (over all choices of $d \leq \Delta$ and arbitrary w_i). Then by the definition of $R(G, v, L)$, for every $L > 0$,

$$T(L) \leq \sum_{i=1}^d \sum_{j=1}^{w_i} T(L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor) + O(n).$$

It is easy to verify that $T(L) = n\Delta^{O(L)} = \left(\frac{n}{\varepsilon}\right)^{O(\log \Delta)}$ for our choice of L . Thus our algorithm is an FPTAS for computing $Z(G)$. \square

B Proof of Lemma 4

Proof. By the definition of R_v , we have

$$R_v = \frac{\Pr_G[v \in I]}{\Pr_G[v \notin I]} = \lambda \cdot \frac{\Pr_{G'}\left[\bigwedge_{i=1}^d v_i \in I\right]}{\Pr_{G'}\left[\bigwedge_{i=1}^d v_i \notin I\right]} = \lambda \cdot \prod_{i=1}^d \frac{\Pr_{G'}\left[v_i \in I \wedge \bigwedge_{j=1}^{i-1} v_j \notin I \wedge \bigwedge_{j=i+1}^d v_j \in I\right]}{\Pr_{G'}\left[v_i \notin I \wedge \bigwedge_{j=1}^{i-1} v_j \notin I \wedge \bigwedge_{j=i+1}^d v_j \in I\right]}.$$

For every $i \in [d]$, define $G_i \triangleq G' - \{v_k \mid i < k \leq d\} - \{e_k \mid 1 \leq k < i\}$, we have

$$\frac{\Pr_{G'}\left[v_i \in I \wedge \bigwedge_{j=1}^{i-1} v_j \notin I \wedge \bigwedge_{j=i+1}^d v_j \in I\right]}{\Pr_{G'}\left[v_i \notin I \wedge \bigwedge_{j=1}^{i-1} v_j \notin I \wedge \bigwedge_{j=i+1}^d v_j \in I\right]} = \frac{\Pr_{G'}\left[v_i \in I \mid \bigwedge_{j=1}^{i-1} v_j \notin I \wedge \bigwedge_{j=i+1}^d v_j \in I\right]}{\Pr_{G'}\left[v_i \notin I \mid \bigwedge_{j=1}^{i-1} v_j \notin I \wedge \bigwedge_{j=i+1}^d v_j \in I\right]} = \frac{\Pr_{G_i}[v_i \in I]}{\Pr_{G_i}[v_i \notin I]}.$$

This is because fixing $v_j \in I$ is equivalent to removing v_j from the graph and fixing $v_j \notin I$ is equivalent to removing all edges incident to v_j from the graph.

Since e_i is the unique hyperedge in G_i that contains v_i , we have

$$\frac{\Pr_{G_i}[v_i \in I]}{\Pr_{G_i}[v_i \notin I]} = 1 - \Pr_{G_i - v_i - e_i}\left[\bigwedge_{j=1}^{w_i} v_{ij} \in I\right] = 1 - \prod_{j=1}^{w_i} \Pr_{G_{ij}}[v_{ij} \in I] = 1 - \prod_{j=1}^{w_i} \frac{R_{ij}}{1 + R_{ij}}.$$

\square

C Proof of Lemma 9

Proof. Let $\lambda'_c \triangleq \frac{d^d}{(d-1)^{d+1}}$ be the uniqueness threshold for the d -ary tree. Then $\lambda < \lambda_c \leq \lambda'_c$.

Plugging $\mathbf{w} = \mathbf{1}$ into (3), we have

$$g_{\mathbf{z}}(\mathbf{1}) = \sqrt{\frac{\lambda \prod_{i=1}^d z_i}{1 + \lambda \prod_{i=1}^d z_i}} \cdot \sum_{i=1}^d \sqrt{1 - z_i}.$$

Let $z = \left(\prod_{i=1}^d z_i\right)^{\frac{1}{d}}$, it follows from Jensen's inequality that

$$g(\mathbf{t}, \mathbf{1}) \leq d \sqrt{\frac{\lambda z^d (1-z)}{1 + \lambda z^d}} < d \sqrt{\frac{\lambda'_c z^d (1-z)}{1 + \lambda'_c z^d}} \quad (5)$$

Recall that $f_{\lambda,d}(x) = \lambda \left(\frac{1}{1+x}\right)^d$. Let \hat{x} be the positive fixed-point of $f_{\lambda'_c,d}(x)$ and $\hat{z} = \frac{1}{1+\hat{x}}$. We show that $d \sqrt{\frac{\lambda'_c z^d (1-z)}{1 + \lambda'_c z^d}}$ achieves its maximum when $z = \hat{z}$. The derivative of $\frac{\lambda'_c z^d (1-z)}{1 + \lambda'_c z^d}$ with respect to z is

$$\left(\frac{\lambda'_c z^d (1-z)}{1 + \lambda'_c z^d}\right)' = -\frac{\lambda'_c z^{d-1}}{(1 + \lambda'_c z^d)^2} (z + \lambda'_c z^{d+1} - d(1-z)).$$

Since $\lambda'_c = \frac{d^d}{(d-1)^{d+1}}$, the above achieves maximum at $\tilde{z} = \frac{d-1}{d}$. If we let $\tilde{x} = \frac{1-\tilde{z}}{\tilde{z}}$, then it is easy to verify that $f_{\lambda'_c,d}(\tilde{x}) = \tilde{x}$, which implies $\hat{z} = \tilde{z}$ because of the uniqueness of the positive fixed-point.

Therefore, we have for some $\alpha < 1$,

$$g_{\mathbf{z}}(\mathbf{1}) \leq \alpha < d \sqrt{\frac{\lambda'_c \hat{z}^d (1-\hat{z})}{1 + \lambda'_c \hat{z}^d}} = |f'_{\lambda'_c,d}(\hat{x})| = 1.$$

□

D FPTAS for Hypergraph Ising Model

D.1 The algorithm and recursion for computing marginal probability

In this section, we give the recursion function and design an algorithm to compute marginal probability in hypergraph Ising Model. Then we prove that the algorithm is indeed an FPTAS for any instance $G(V, \mathcal{E}, (\beta, \gamma))$ if $1 - \frac{2}{2e^{-1/2}\Delta+3} \leq \beta(e), \gamma(e) \leq 1$.

Graph Operations We first fix some notations on graph modification specific to our hypergraph two state spin model. Let $G(V, \mathcal{E}, (\beta, \gamma))$ be an instance. The first groups of operations are about vertex removal and edge removal, which is similar to the case of hardcore model:

- For every $v \in V$, we denote $G - v \triangleq (V \setminus \{v\}, \mathcal{E}', (\beta', \gamma'))$ where $\mathcal{E}' \triangleq \{e - v \mid e \in \mathcal{E}\}$ and $\beta'(e - v) = \beta(e)$, $\gamma'(e - v) = \gamma(e)$ for every $e \in \mathcal{E}$.
- For every $e \in \mathcal{E}$, we denote $G - e \triangleq (V, \mathcal{E} \setminus \{e\}, (\beta, \gamma))$.
- Let x be a vertex or an edge and y be a vertex or an edge, we denote $G - x - y \triangleq (G - x) - y$.
- Let $S = \{v_1, \dots, v_k\} \subseteq V$, we denote $G - S \triangleq G - v_1 - v_2 \cdots - v_k$.
- Let $\mathcal{F} = \{e_1, \dots, e_k\} \subseteq \mathcal{E}$, we denote $G - \mathcal{F} \triangleq G - e_1 - e_2 \cdots - e_k$.

The second group of operations is about *pinning* the value of a vertex, whose effect is to change $\beta(e)$ or $\gamma(e)$ for edge e incident to it.

- Let $v \in V$ be a vertex, we denote $G|_{v=0} \triangleq (V \setminus \{v\}, \mathcal{E}', (\beta', \gamma')_{e \in \mathcal{E}'})$ where $\mathcal{E}' \triangleq \{e - v \mid e \in \mathcal{E}\}$, $\beta'(e - v) \triangleq \beta(e)$ for every $e \in \mathcal{E}$ and

$$\gamma'(e - v) \triangleq \begin{cases} \gamma(e) & \text{if } v \notin e \\ 1 & \text{otherwise.} \end{cases}$$

This operation is to pin the value of v to 0.

- Similarly, for a vertex $v \in V$, we denote $G|_{v=1} \triangleq (V \setminus \{v\}, \mathcal{E}', (\beta', \gamma'))$ where $\mathcal{E}' \triangleq \{e \setminus \{v\} \mid e \in \mathcal{E}\}$, $\gamma'(e - v) \triangleq \gamma(e)$ for every $e \in \mathcal{E}$ and

$$\beta'(e - v) \triangleq \begin{cases} \beta(e) & \text{if } v \notin e \\ 1 & \text{otherwise.} \end{cases}$$

This operation is to pin the value of v to 1.

- Let $u, v \in V$ be two vertices and $i, j \in \{0, 1\}$. We denote $G|_{u=i, v=j} \triangleq (G|_{u=i})|_{v=j}$ and this notation generalizes to more vertices.
- Let $S = \{v_1, \dots, v_k\} \subseteq V$, we use $G|_{S=0}$ (resp. $G|_{S=1}$) to denote $G|_{v_1=0, v_2=0, \dots, v_k=0}$ and $G|_{v_1=1, v_2=1, \dots, v_k=1}$.

Let $G(V, \mathcal{E}, (\beta, \gamma))$ be an instance and $v \in V$ be an arbitrary vertex with degree d . Let $E(v) = \{e_1, \dots, e_d\}$ be the set of edges incident to v . For every $i \in [d]$, we assume $e_i = \{v\} \cup \{v_{ij} \mid j \in [w_i]\}$ consists of $w_i + 1$ vertices where $w_i \geq 0$.

We define a new instance $G'(V', \mathcal{E}', (\beta'_e, \gamma'_e))$ which is obtained from G by replacing v by d copies of itself and each e_i contains a distinct copy. Formally, $V' \triangleq (V \setminus \{v\}) \cup \{v_1, \dots, v_d\}$ and $\mathcal{E}' \triangleq (\mathcal{E} \setminus E(v)) \cup \{e_i - v + v_i \mid i \in [d]\}$; $\beta'(e) = \beta(e)$, $\gamma'(e) = \gamma(e)$ for every $e \in \mathcal{E} \setminus E(v)$ and $\beta'(e_i - v + v_i) = \beta(e_i)$, $\gamma'(e_i - v + v_i) = \gamma(e_i)$ for every $i \in [d]$.

For every $i \in [d]$ and $j \in [w_i]$, we define two smaller instances $G_{ij}^0(V_{ij}^0, \mathcal{E}_{ij}^0, (\beta_{ij}^0, \gamma_{ij}^0))$ and $G_{ij}^1(V_{ij}^1, \mathcal{E}_{ij}^1, (\beta_{ij}^1, \gamma_{ij}^1))$:

- $G_{ij}^0 \triangleq (((G'|_{V_{<i}=0})|_{V_{>i}=1}) - e_i)|_{V_{<j}^i=0}$;
- $G_{ij}^1 \triangleq (((G'|_{V_{<i}=0})|_{V_{>i}=1}) - e_i)|_{V_{<j}^i=1}$,

where $V_{<i} \triangleq \{v_k \mid k < i\}$, $V_{>i} \triangleq \{v_k \mid k > i\}$, $V_{<j}^i \triangleq \{v_{ik} \mid k < j\}$.

We now define the ratio of the probability on instances defined above.

- Let $R_v \triangleq \frac{\Pr_G[\sigma(v)=1]}{\Pr_G[\sigma(v)=0]}$;
- For every $i \in [d]$ and $j \in [w_i]$, let $R_{ij}^0 \triangleq \frac{\Pr_{G_{ij}^0}[\sigma(v_{ij})=1]}{\Pr_{G_{ij}^0}[\sigma(v_{ij})=0]}$, $R_{ij}^1 \triangleq \frac{\Pr_{G_{ij}^1}[\sigma(v_{ij})=1]}{\Pr_{G_{ij}^1}[\sigma(v_{ij})=0]}$.

It follows from our definition that for every $i \in [d]$, $R_{i1}^0 = R_{i1}^1$.

We can compute R_v by the following recursion:

Lemma 10.

$$R_v = \lambda \cdot \prod_{i=1}^d \frac{1 - (1 - \gamma(e_i)) \frac{R_{i1}^0}{1 + R_{i1}^0} \prod_{j=2}^{w_i} \frac{R_{ij}^1}{1 + R_{ij}^1}}{1 - (1 - \beta(e_i)) \frac{1}{1 + R_{i1}^0} \prod_{j=2}^{w_i} \frac{1}{1 + R_{ij}^0}}. \quad (6)$$

Proof. By the definition of R_v , we have

$$R_v = \frac{\Pr_G[\sigma(v) = 1]}{\Pr_G[\sigma(v) = 0]} = \lambda \cdot \frac{\Pr_{G'}\left[\bigwedge_{i=1}^d \sigma(v_i) = 1\right]}{\Pr_{G'}\left[\bigwedge_{i=1}^d \sigma(v_i) = 0\right]} = \lambda \cdot \prod_{i=1}^d \frac{\Pr_{G'}\left[\sigma(v_i) = 1 \wedge \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0 \wedge \bigwedge_{j=i+1}^d \sigma(v_j) = 1\right]}{\Pr_{G'}\left[\sigma(v_i) = 0 \wedge \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0 \wedge \bigwedge_{j=i+1}^d \sigma(v_j) = 1\right]}.$$

For every $i \in [d]$, define $G_i \triangleq (G'|_{V_{<i}=\mathbf{0}})|_{V_{>i}=\mathbf{1}}$, then

$$\begin{aligned} \frac{\Pr_{G'}\left[\sigma(v_i) = 1 \wedge \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0 \wedge \bigwedge_{j=i+1}^d \sigma(v_j) = 1\right]}{\Pr_{G'}\left[\sigma(v_i) = 0 \wedge \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0 \wedge \bigwedge_{j=i+1}^d \sigma(v_j) = 1\right]} &= \frac{\Pr_{G'}\left[\sigma(v_i) = 1 \mid \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0 \wedge \bigwedge_{j=i+1}^d \sigma(v_j) = 1\right]}{\Pr_{G'}\left[\sigma(v_i) = 0 \mid \bigwedge_{j=1}^{i-1} \sigma(v_j) = 0 \wedge \bigwedge_{j=i+1}^d \sigma(v_j) = 1\right]} \\ &= \frac{\Pr_{G_i}[\sigma(v_i) = 1]}{\Pr_{G_i}[\sigma(v_i) = 0]}. \end{aligned}$$

Since e_i is the unique edge in G_i that contains v_i , we have

$$\begin{aligned} \frac{\Pr_{G_i}[\sigma(v_i) = 1]}{\Pr_{G_i}[\sigma(v_i) = 0]} &= \frac{\gamma(e_i)\Pr_{G_i-v_i-e_i}\left[\bigwedge_{j=1}^{w_i} \sigma(v_{ij}) = 1\right] + \left(1 - \Pr_{G_i-v_i-e_i}\left[\bigwedge_{j=1}^{w_i} \sigma(v_{ij}) = 1\right]\right)}{\beta(e_i)\Pr_{G_i-v_i-e_i}\left[\bigwedge_{j=1}^{w_i} \sigma(v_{ij}) = 0\right] + \left(1 - \Pr_{G_i-v_i-e_i}\left[\bigwedge_{j=1}^{w_i} \sigma(v_{ij}) = 0\right]\right)} \\ &= \frac{1 - (1 - \gamma(e_i))\Pr_{G_i-v_i-e_i}[\sigma(v_{i1}) = 1] \cdot \prod_{j=2}^{w_i} \Pr_{G_i-v_i-e_i}\left[\sigma(v_{ij}) = 1 \mid \bigwedge_{k=1}^{j-1} \sigma(v_k) = 1\right]}{1 - (1 - \beta(e_i))\Pr_{G_i-v_i-e_i}[\sigma(v_{i1}) = 0] \cdot \prod_{j=2}^{w_i} \Pr_{G_i-v_i-e_i}\left[\sigma(v_{ij}) = 0 \mid \bigwedge_{k=1}^{j-1} \sigma(v_k) = 0\right]} \\ &= \frac{1 - (1 - \gamma(e_i))\Pr_{G_{i1}^1}[\sigma(v_{i1}) = 1] \cdot \prod_{j=2}^{w_i} \Pr_{G_{ij}^1}[\sigma(v_{ij}) = 1]}{1 - (1 - \beta(e_i))\Pr_{G_{i1}^0}[\sigma(v_{i1}) = 0] \cdot \prod_{j=2}^{w_i} \Pr_{G_{ij}^0}[\sigma(v_{ij}) = 0]} \\ &= \frac{1 - (1 - \gamma(e_i))\frac{R_{i1}^0}{1+R_{i1}^0} \prod_{j=2}^{w_i} \frac{R_{ij}^1}{1+R_{ij}^1}}{1 - (1 - \beta(e_i))\frac{1}{1+R_{i1}^0} \prod_{j=2}^{w_i} \frac{1}{1+R_{ij}^0}}. \end{aligned}$$

The last equality is due to the fact that $R_{i1}^0 = R_{i1}^1$. □

The algorithm description is similar to Section 3.2. Let $G(V, \mathcal{E}, (\beta, \gamma))$ be an instance of our two state spin model with maximum degree Δ , and $v \in V$ be an arbitrary vertex with degree d . Let $E(v) = \{e_1, \dots, e_d\}$ denote the set of edges incident to v . Define $G_{ij}^0, G_{ij}^1, R_v, R_{ij}^0, R_{ij}^1$ as in Section D. Then the recursion (6) gives a way to compute the marginal probability $\Pr_G[\sigma(v) = 1]$ *exactly*. However, an exact evaluation of the recursion requires a computation tree with exponential size. Thus we introduce the following truncated version of the recursion, with respect to constants $c > 0$ and $0 < \alpha < 1$. Let

$$f_G(\mathbf{r}) = \lambda \cdot \prod_{i=1}^d \frac{1 - (1 - \gamma(e_i))\frac{r_{i1}^0}{1+r_{i1}^0} \prod_{j=2}^{w_i} \frac{r_{ij}^1}{1+r_{ij}^1}}{1 - (1 - \beta(e_i))\frac{1}{1+r_{i1}^0} \prod_{j=2}^{w_i} \frac{1}{1+r_{ij}^0}},$$

where $\mathbf{r} \triangleq ((r_{ij}^0)_{1 \leq j \leq w_i}, (r_{ij}^1)_{2 \leq j \leq w_i})_{i \in [d]}$ with

$$\begin{aligned} r_{ij}^0 &= R(G_{ij}^0, v_{ij}, L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor) \\ r_{ij}^1 &= R(G_{ij}^1, v_{ij}, L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor). \end{aligned}$$

We can describe our truncated recursion as follows:

$$R(G, v, L) = \begin{cases} f_G(\mathbf{r}) & \text{if } L > 0 \\ \lambda & \text{otherwise.} \end{cases}$$

The recursion can be directly used to compute $R(G, v, L)$ for any given L and it induces a truncated computation tree (with height L in some special metric).

We claim that $R(G, v, L)$ is a good estimate of R_v for a suitable choice of c and α , for those (β, γ, Δ) satisfying that $1 - \frac{2}{2e^{-1/2}\Delta+3} \leq \beta(e), \gamma(e) \leq 1$ for all $e \in \mathcal{E}$.

Lemma 11. *Let $G(V, \mathcal{E}, (\beta, \gamma))$ be an instance of our generalized two state spin system model with maximum degree $\Delta \geq 2$. Let $v \in V$ be a vertex and assume $1 - \frac{2}{2e^{-1/2}\Delta+3} \leq \beta(e), \gamma(e) \leq 1$ for all $e \in \mathcal{E}$. There exist constants $C > 0$ and $0 < \alpha < 1$ such that*

$$|R(G, v, L) - R_v| \leq C \cdot \alpha^{\max\{0, L\}}$$

for every L .

The Lemma 11 will be proved in the next section. The proof of Theorem 3 with this lemma is almost identical to the proof of Theorem 1 and we omit it here.

D.2 Correlation decay

Again, the key is to bound the amortized decay rate as in the following lemma.

Lemma 12. *Let $\Delta \geq 2$ be a constant integer. Suppose $d \leq \Delta$ is a constant integer, $\lambda \in (0, 1)$, $\beta_c = 1 - \frac{2}{2e^{-1/2}\Delta+3}$ are constant real numbers and $\gamma_1, \gamma_2, \dots, \gamma_d, \beta_1, \beta_2, \dots, \beta_d \in [\beta_c, 1]$. Let $w_1, \dots, w_d > 0$ be integers and*

$$f(\mathbf{r}) = \lambda \prod_{i=1}^d \frac{1 - (1 - \gamma_i) \frac{r_{i1}^0}{1+r_{i1}^0} \prod_{j=2}^{w_i} \frac{r_{ij}^1}{1+r_{ij}^1}}{1 - (1 - \beta_i) \frac{1}{1+r_{i1}^0} \prod_{j=2}^{w_i} \frac{1}{1+r_{ij}^0}}$$

be a $\left(\sum_{i=1}^d w_i\right)$ -ary function. Let $\Phi(x) = \frac{1}{x}$. There exist constants $\alpha < 1$ and $c > 0$ depending on γ and Δ (but not depending on w_i) such that

$$\sum_{i=1}^d w_i^c \left(\sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^0)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^0} \right| + \sum_{j=2}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^1)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^1} \right| \right) \leq \alpha < 1$$

for every $\mathbf{r} = (r_{ij})_{i \in [d], j \in [w_i]}$ where each $r_{ij} \in [\lambda\beta_c^\Delta, \lambda\beta_c^{-\Delta}]$.

Proof. We first let $x_i \triangleq \frac{r_{i1}^0}{1+r_{i1}^0}$, $y_{ij} \triangleq \frac{r_{ij}^1}{1+r_{ij}^1}$, $z_{ij} \triangleq \frac{r_{ij}^0}{1+r_{ij}^0}$ for all $1 \leq i \leq d$ and $2 \leq j \leq w_i$. Then it holds that $x_i, y_{ij}, z_{ij} \in \left[\frac{\lambda\beta_c^\Delta}{1+\lambda\beta_c^\Delta}, \frac{\lambda}{1+\lambda\beta_c^\Delta} \right]$. To simplify the notation, we denote

$$A_i \triangleq 1 - (1 - \gamma_i) x_i \prod_{j=2}^{w_i} y_{ij}$$

$$B_i \triangleq 1 - (1 - \beta_i) (1 - x_i) \prod_{j=2}^{w_i} (1 - z_{ij})$$

where $A_i \in [\gamma_i, 1] \subseteq [\beta_c, 1]$, $B_i \in [\beta_i, 1] \subseteq [\beta_c, 1]$.

Then we can write f as

$$f = \lambda \prod_{i=1}^d \frac{A_i}{B_i}.$$

We can directly compute the partial derivatives of f , which yields the following for all $i \in [d]$ and $2 \leq j \leq w_i$:

$$\begin{aligned} \frac{\partial f}{\partial r_{i1}^0} &= \left(\lambda \prod_{\substack{k \in [d] \\ k \neq i}} \frac{A_k}{B_k} \right) \cdot \frac{(1-x_i)((1-A_i)(1-B_i) - (1-x_i)(1-A_i) - x_i(1-B_i))}{x_i B_i^2} \\ &= f \cdot \frac{(1-x_i)((1-A_i)(1-B_i) - (1-x_i)(1-A_i) - x_i(1-B_i))}{x_i A_i B_i}; \\ \frac{\partial f}{\partial r_{ij}^1} &= \left(\lambda \prod_{\substack{k \in [d] \\ k \neq i}} \frac{A_k}{B_k} \right) \cdot \frac{-(1-y_{ij})^2(1-A_i)}{y_{ij} B_i} = f \cdot \frac{-(1-y_{ij})^2(1-A_i)}{y_{ij} A_i}; \\ \frac{\partial f}{\partial r_{ij}^0} &= \left(\lambda \prod_{\substack{k \in [d] \\ k \neq i}} \frac{A_k}{B_k} \right) \cdot \frac{-A_i(1-B_i)(1-z_{ij})}{B_i^2} = f \cdot \frac{-(1-B_i)(1-z_{ij})}{B_i}. \end{aligned}$$

Thus,

$$\begin{aligned} & \sum_{i=1}^d w_i^c \left(\sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^0)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^0} \right| + \sum_{j=2}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^1)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^1} \right| \right) \\ &= \frac{1}{f(\mathbf{r})} \sum_{i=1}^d w_i^c \left(\frac{x_i}{1-x_i} \left| \frac{\partial f(\mathbf{r})}{\partial r_{i1}^0} \right| + \sum_{j=2}^{w_i} \frac{y_i}{1-y_i} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^1} \right| + \sum_{j=2}^{w_i} \frac{z_i}{1-z_i} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^0} \right| \right) \\ &= \sum_{i=1}^d w_i^c \left(\frac{-(1-A_i)(1-B_i) + (1-x_i)(1-A_i) + x_i(1-B_i)}{A_i B_i} + \frac{1-A_i}{A_i} \sum_{j=2}^{w_i} (1-y_{ij}) + \frac{1-B_i}{B_i} \sum_{j=2}^{w_i} z_{ij} \right) \end{aligned}$$

Let $y_i \triangleq \left(\prod_{j=2}^{w_i} y_{ij} \right)^{1/(w_i-1)}$, $z_i \triangleq 1 - \left(\prod_{j=2}^{w_i} (1-z_{ij}) \right)^{1/(w_i-1)}$ for every $i \in [d]$, then it holds that

$$\begin{aligned} A_i &= 1 - (1-\gamma_i)x_i y_i^{w_i-1}, \\ B_i &= 1 - (1-\beta_i)(1-x_i)(1-z_i)^{w_i-1}, \end{aligned}$$

$$\begin{aligned} \sum_{j=2}^{w_i} (1-y_{ij}) &\leq (w_i-1)(1-y_i) \\ \sum_{j=2}^{w_i} z_{ij} &\leq (w_i-1)z_i. \end{aligned}$$

Since $A_i, B_i \in [\beta_c, 1]$ for every $i \in [d]$, we have

$$\begin{aligned}
& \sum_{i=1}^d w_i^c \left(\frac{-(1-A_i)(1-B_i) + (1-x_i)(1-A_i) + x_i(1-B_i)}{A_i B_i} + \frac{1-A_i}{A_i} \sum_{j=2}^{w_i} (1-y_{ij}) + \frac{1-B_i}{B_i} \sum_{j=2}^{w_i} z_{ij} \right) \\
& \leq \sum_{i=1}^d w_i^c \left(\frac{(1-x_i)(1-A_i) + x_i(1-B_i) - (1-A_i)(1-B_i)}{A_i B_i} + (w_i-1) \left(\frac{(1-A_i)(1-y_i)}{A_i} + \frac{(1-B_i)z_i}{B_i} \right) \right) \\
& \leq \sum_{i=1}^d w_i^c \left(\frac{(1-x_i)(1-A_i) + x_i(1-B_i)}{A_i B_i} + (1-\beta_c)(w_i-1) \left(\frac{x_i y_i^{w_i-1} (1-y_i)}{A_i} + \frac{(1-x_i)(1-z_i)^{w_i-1} z_i}{B_i} \right) \right) \\
& \leq \frac{1-\beta_c}{\beta_c^2} \sum_{i=1}^d w_i^c \left((1-x_i)x_i(y_i^{w_i-1} + (1-z_i)^{w_i-1}) + \beta_c(w_i-1)(x_i y_i^{w_i-1} (1-y_i) + (1-x_i)(1-z_i)^{w_i-1} z_i) \right)
\end{aligned}$$

We now assume Lemma 13, thus there exist constants $\alpha < 1, c > 0$ depending on λ and Δ such that

$$w_i^c \left((1-x_i)x_i(y_i^{w_i-1} + (1-z_i)^{w_i-1}) + \beta_c(w_i-1)(x_i y_i^{w_i-1} (1-y_i) + (1-x_i)(1-z_i)^{w_i-1} z_i) \right) \leq \alpha \beta_c e^{\frac{1}{2\beta_c}-1}$$

for every $i \in [d]$. Thus,

$$\begin{aligned}
& \sum_{i=1}^d w_i^c \left(\frac{-(1-A_i)(1-B_i) + (1-x_i)(1-A_i) + x_i(1-B_i)}{A_i B_i} + \frac{1-A_i}{A_i} \sum_{j=2}^{w_i} (1-y_{ij}) + \frac{1-B_i}{B_i} \sum_{j=2}^{w_i} z_{ij} \right) \\
& \leq \frac{1-\beta_c}{\beta_c} \cdot d \cdot \alpha \cdot e^{\beta_c^{-1}/2-1} \\
& \leq \frac{2\alpha d}{2e^{-1/2}d+3} \cdot \frac{2e^{-1/2}d+3}{2e^{-1/2}d+1} \cdot e^{(1-2/(2e^{-1/2}d+3))^{-1}/2-1} \\
& = \alpha \cdot \frac{2d}{2e^{-1/2}d+1} \cdot e^{(1-2/(2e^{-1/2}d+3))^{-1}/2-1} \\
& \leq \alpha.
\end{aligned}$$

The last inequality is due to the fact that $h(d) \triangleq \frac{2d}{2e^{-1/2}d+1} \cdot e^{(1-2/(2e^{-1/2}d+3))^{-1}/2-1}$ is an increasing function on d and $\lim_{d \rightarrow \infty} h(d) = 1$.

Combining all above, there exist constants $\alpha < 1$ and $c > 0$ such that for all $w_1, w_2, \dots, w_d > 0$ and every $\mathbf{r} = (r_{ij})_{i \in [d], j \in [w_i]}$ with each $r_{ij} \in [\lambda \beta_c^\Delta, \lambda \beta_c^{-\Delta}]$, it holds

$$\sum_{i=1}^d w_i^c \left(\sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^0)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^0} \right| + \sum_{j=2}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^1)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^1} \right| \right) \leq \alpha < 1$$

□

It remains to prove Lemma 13:

Lemma 13. *For all $\beta < 1$ and $0 < \delta < 1/2$, there exist constants $\alpha < 1$ and $c > 0$ such that*

$$w^c \left((1-x)x(y^{w-1} + (1-z)^{w-1}) + \beta(w-1)xy^{w-1}(1-y) + \beta(w-1)(1-x)(1-z)^{w-1}z \right) \leq \alpha \beta e^{\frac{1}{2\beta}-1}$$

for all $w \in \mathbb{N}$ and all $x, y, z \in [\delta, 1-\delta]$.

Proof. For a fixed w , let

$$V_w(x, y, z) \triangleq (1-x)x(y^{w-1} + (1-z)^{w-1}) + \beta(w-1)xy^{w-1}(1-y) + \beta(w-1)(1-x)(1-z)^{w-1}z.$$

Then $V_w(x, y, z)$ achieves maximum value when

$$y = y_c \triangleq \frac{\beta(w-1) + 1 - x}{\beta w} = 1 - \frac{1 - (1-x)/\beta}{w},$$

$$z = z_c \triangleq 1 - \frac{\beta(w-1) + x}{\beta w} = \frac{1 - x/\beta}{w}.$$

If $y < 1 - \frac{1-(1-x)/\beta}{w}$, $V_w(x, y, z)$ is an increasing function on y . If $z > \frac{1-x/\beta}{w}$, $V_w(x, y, z)$ is a decreasing function on z .

Let $\hat{w} \triangleq \max \left\{ \left\lceil \frac{1}{\delta} \right\rceil, \left\lceil -\frac{1}{\ln(1-\delta)} \right\rceil \right\} + 1$. If $w > \hat{w}$, the monotonicity of $V_w(x, y, z)$ implies that the function achieves its maximum when the y and z are at the boundary, i.e., $y = 1 - z = 1 - \delta$, for every fixed x .

We now analyze the case $w \leq \hat{w}$ and the case $w > \hat{w}$ separately. Define $\alpha \triangleq \left(\frac{(1 + \frac{1/(2\beta)-1}{\hat{w}})^{\hat{w}}}{e^{1/(2\beta)-1}} \right)^{1/2}$.

It is easy to verify that $\alpha < 1$ since $(1 + x/n)^n < e^x$ for every $x > 0$.

- (If $w \leq \hat{w}$.) We plug $y = y_c$ and $z = z_c$ into $V_w(x, y, z)$:

$$V_w(x, y, z) \leq (1-x)x(y_c^{w-1} + (1-z_c)^{w-1}) + \beta(w-1)xy_c^{w-1}(1-y_c) + \beta(w-1)(1-x)(1-z_c)^{w-1}z_c.$$

The derivative of the right hand side with respect to x is

$$\frac{(\beta(w-1) + 1 - x)^{w-1}(1 + \beta(w-1) - (w+1)x) - (\beta(w-1) + x)^{w-1}(1 + \beta(w-1) - (w+1)(1-x))}{(\beta w)^w}.$$

It is easy to see that the function above is zero when $x = 1/2$ and positive for smaller x , negative for larger x .

Thus, when $x = 1/2, y = y_c, z = z_c$, $V_w(x, y, z)$ achieves its maximum value

$$V_w\left(\frac{1}{2}, 1 - \frac{1 - \frac{1}{2\beta}}{w}, \frac{1 - \frac{1}{2\beta}}{w}\right) = \beta \left(1 - \frac{1 - \frac{1}{2\beta}}{w}\right)^w \leq \beta \left(1 - \frac{1 - \frac{1}{2\beta}}{\hat{w}}\right)^{\hat{w}} < \beta e^{\frac{1}{2\beta}-1}.$$

Let $c_1 \triangleq \log_{\hat{w}} \frac{\alpha e^{\frac{1}{2\beta}-1}}{\left(1 - \frac{1 - \frac{1}{2\beta}}{\hat{w}}\right)^{\hat{w}}} = -\log_{\hat{w}} \alpha > 0$, then for all $w \leq \hat{w}$ and $c' \in (0, c_1]$,

$$w^{c'} V_w(x, y, z) \leq w^{c_1} V_w(x, y, z) \leq \hat{w}^{c_1} \beta \left(1 - \frac{1 - \frac{1}{2\beta}}{\hat{w}}\right)^{\hat{w}} = \frac{\beta}{\alpha} \left(1 - \frac{1 - \frac{1}{2\beta}}{\hat{w}}\right)^{\hat{w}} = \alpha \beta e^{\frac{1}{2\beta}-1}.$$

- (If $w > \hat{w}$.) Let $c_2 \triangleq -\hat{w} \log(1 - \delta) - 1 > 0$. Then

$$V_w(x, y, z) \leq V_w(x, 1 - \delta, \delta) = 2x(1-x)(1-\delta)^{w-1} + \beta(w-1)\delta(1-\delta)^{w-1}$$

The above achieves its maximum when $x = 1/2$, i.e., $V_w(x, y, z) \leq V_w(\frac{1}{2}, 1 - \delta, \delta)$. Therefore

$$w^{c_2} \cdot V_w(x, y, z) \leq w^{c_2} \cdot V_w\left(\frac{1}{2}, 1 - \delta, \delta\right) = w^{c_2} \cdot \frac{1 + 2\beta(w-1)\delta}{2}(1-\delta)^{w-1}.$$

We now prove that for all $c' \in (0, c_2]$, $g(w) \triangleq w^{c' \frac{1+2\beta(w-1)\delta}{2}}(1-\delta)^{w-1}$ is decreasing on w when $w > \hat{w}$. Since $c' \leq c_2 = -\hat{w} \log(1-\delta) - 1 \leq -(w-1) \log(1-\delta) - 1$ for all $w > \hat{w}$,

$$\begin{aligned} \frac{\partial g}{\partial w} &= \frac{w^{c'-1}(1-\delta)^{w-1}}{2} (c' + w \log(1-\delta)) + \beta \delta w^{c'-1}(1-\delta)^{w-1} (w + (w-1)(c' + w \log(1-\delta))) \\ &\leq \frac{w^{c'-1}(1-\delta)^{w-1}}{2} (\log(1-\delta) - 1) + \beta \delta w^{c'-1}(1-\delta)^{w-1} (w + (w-1)(\log(1-\delta) - 1)) \\ &\leq -\frac{w^{c'-1}(1-\delta)^{w-1}}{2} + \beta \delta w^{c'-1}(1-\delta)^{w-1} (1 + (w-1) \log(1-\delta)) \\ &\leq -\frac{w^{c'-1}(1-\delta)^{w-1}}{2} - \beta \delta w^{c'-1}(1-\delta)^{w-1} c' \\ &< 0 \end{aligned}$$

Thus, $g(w)$ is decreasing on w when $w > \hat{w}$. In light of this, $w^{c'} \cdot V_w(x, y, z)$ can be upper bounded by $\hat{w}^{c'} \cdot V_{\hat{w}}(\frac{1}{2}, 1-\delta, \delta)$ for all $c' \in (0, c_2]$.

In all, we have

$$\begin{aligned} w^{c'} \cdot V_w(x, y, z) &\leq w^{c'} \cdot V_w\left(\frac{1}{2}, 1-\delta, \delta\right) \\ &\leq \hat{w}^{c'} \cdot V_{\hat{w}}\left(\frac{1}{2}, 1-\delta, \delta\right) \\ &\leq \hat{w}^{c_1} \cdot V_{\hat{w}}\left(\frac{1}{2}, 1 - \frac{1 - \frac{1}{2\beta}}{\hat{w}}, \frac{1 - \frac{1}{2\beta}}{\hat{w}}\right) \hat{w}^{c'-c_1} \\ &= \alpha \beta e^{\frac{1}{2\beta}-1} \hat{w}^{c'-c_1} \end{aligned}$$

for all $w > \hat{w}$ and $c' \in (0, c_2]$.

Let $c = \min\{c_1, c_2\}$, then we have $w^c V_w(x, y, z) \leq w^{c_1} V_w(x, y, z) \leq \alpha \beta e^{\frac{1}{2\beta}-1}$ if $w \leq \hat{w}$ and $w^c V_w(x, y, z) \leq \hat{w}^{c_1} V_{\hat{w}}(x, y, z) \hat{w}^{c'-c_1} \leq \alpha \beta e^{\frac{1}{2\beta}-1}$ if $w > \hat{w}$, i.e.,

$$w^c ((1-x)x(y^{w-1} + (1-z)^{w-1}) + \beta(w-1)xy^{w-1}(1-y) + \beta(w-1)(1-x)(1-z)^{w-1}z) \leq \alpha \beta e^{\frac{1}{2\beta}-1}$$

for all $w \in \mathbb{N}$ and all $x, y, z \in [\delta, 1-\delta]$. □

We are now ready to prove Lemma 11.

Proof of Lemma 11. Let $\Phi(x) = \frac{1}{x}$ and $\phi(x) = \int \Phi(x) dx = \ln x$. We apply induction on $\ell \triangleq \max\{0, L\}$ to show that, if $d \leq \Delta$, then $|\phi(f_v) - \phi(f(G, v, L))| \leq 4\sqrt{e}\alpha^L$ for some constant $\alpha < 1$.

If $\ell = 0$, note that $f_v \in [\lambda\beta_c^\Delta, \lambda\beta_c^{-\Delta}]$, thus $|\phi(f(G, v, L)) - \phi(f_v)| \leq -2\Delta \ln \beta_c$. Since $|\ln(1-x)| \leq 2x$ for all $x \in (0, \frac{1}{2})$,

$$|\phi(f(G, v, L)) - \phi(f_v)| \leq -2\Delta \ln \beta_c \leq 4\Delta \frac{2}{\frac{2}{\sqrt{e}}\Delta + 3} \leq 4\sqrt{e}.$$

We now assume $L = \ell > 0$ and the lemma holds for smaller ℓ . For every $i \in [d]$ and $j \in [w_i]$, we denote $r_{ij}^0 = R_{ij}^0, r_{ij}^1 = R_{ij}^1$ and $\hat{r}_{ij}^0 = R(G_{ij}^0, v_{ij}, L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor), \hat{r}_{ij}^1 = R(G_{ij}^1, v_{ij}, L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor)$. Let $\mathbf{r} = ((r_{ij}^0)_{1 \leq j \leq w_i}, (r_{ij}^1)_{2 \leq j \leq w_i})_{i \in [d]}$, $\hat{\mathbf{r}} = ((\hat{r}_{ij}^0)_{1 \leq j \leq w_i}, (\hat{r}_{ij}^1)_{2 \leq j \leq w_i})_{i \in [d]}$. Let $f(\mathbf{r}) =$

$\lambda \prod_{i=1}^d \frac{1-(1-\gamma_i) \frac{r_{i1}^0}{1+r_{i1}^0} \prod_{j=2}^{w_i} \frac{r_{ij}^1}{1+r_{ij}^1}}{1-(1-\beta_i) \frac{1}{1+r_{i1}^0} \prod_{j=2}^{w_i} \frac{1}{1+r_{ij}^0}}$, then it follows from Proposition 7 that for some $\tilde{\mathbf{r}} = ((\tilde{r}_{ij}^0)_{1 \leq j \leq w_i}, (\tilde{r}_{ij}^1)_{2 \leq j \leq w_i})_{i \in [d]}$ where each $\tilde{r}_{ij}^0, \tilde{r}_{ij}^1 \in [\lambda\beta_c^\Delta, \lambda\beta_c^{-\Delta}]$,

$$\begin{aligned} |\phi(R_v) - \phi(R(G, v, L))| &\leq \sum_{i=1}^d \left(\sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^0)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^0} \right| \cdot |\phi(r_{ij}^0) - \phi(\hat{r}_{ij}^0)| + \sum_{j=2}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^1)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^1} \right| \cdot |\phi(r_{ij}^1) - \phi(\hat{r}_{ij}^1)| \right) \\ &\stackrel{(\spadesuit)}{\leq} 4\sqrt{e} \sum_{i=1}^d \left(\sum_{j=1}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^0)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^0} \right| + \sum_{j=2}^{w_i} \frac{\Phi(f)}{\Phi(r_{ij}^1)} \left| \frac{\partial f(\mathbf{r})}{\partial r_{ij}^1} \right| \right) \alpha^{L - \lfloor 1 + c \log_{1/\alpha} w_i \rfloor} \\ &\stackrel{(\heartsuit)}{\leq} 4\sqrt{e} \alpha^L. \end{aligned}$$

(\spadesuit) follows from the induction hypothesis and (\heartsuit) is due to Lemma 12 for all $d \leq \Delta$.

Then the lemma follows from Proposition 7, since

$$\begin{aligned} |R_v - R(G, v, L)| &= \frac{1}{\Phi(\tilde{x})} \cdot |\phi(R_v) - \phi(R(G, v, L))| \quad \text{for some } \tilde{x} \in [\lambda\beta_c^\Delta, \lambda\beta_c^{-\Delta}] \\ &\leq e^{\sqrt{e}} \lambda^{-1} \cdot 4\sqrt{e} \alpha^L \\ &= 4\lambda^{-1} e^{\frac{1}{2} + \sqrt{e}} \alpha^L. \end{aligned}$$

The inequality above is due to $\beta_c^{-\Delta} = \left(1 - \frac{\sqrt{e}}{\Delta + \frac{3}{2}\sqrt{e}}\right)^{-\Delta} \leq e^{\sqrt{e}}$. □